Supplementary Materials

Materials and Methods

Patients

There were 16 sessions with 5 patients in total (**Table S1**). All participants provided written informed consent using procedures approved by the Internal Review Board of West Virginia University (WVU).

Stimuli

We used faces of celebrities from the CelebA dataset (1). We selected 50 identities with 10 images for each identity, totaling 500 face images. The identities were selected according to gender and race as well as facial landmarks related to face similarity (e.g., oval face and pointy nose). We used the same stimuli for all patients. Patients were asked to indicate whether they were familiar with each identity in a follow-up survey.

We further used two validation datasets. First, we used a newly-collected FBI Twins Dataset that included pairs of colored photos with the following relationships: identical twins (IT), mirror twins (MT), fraternal twins (FT), mother-child (MC), father-child (FC), and spouses (SP). Therefore, this dataset contained faces with various levels of similarity, and all faces from this dataset were unfamiliar to the patients. The photographing conditions were well controlled to ensure similar background and lighting, and all photos are in high resolution (3840 × 5760).

Second, we used a FaceGen Dataset with model faces, which notably contained only feature information but no real identity information. We used the FaceGen Modeller program (http://facegen.com; version 3.1) to randomly generate 300 faces (see (2) for detailed procedures). FaceGen constructs face space models using information extracted from 3D laser scans of real faces. To create the face space model, the shape of a face was represented by the vertex positions of a polygonal model of fixed mesh topology. With the vertex positions, a principal component

analysis (PCA) was used to extract the components that accounted for most of the variance in face shape. Each principal component (PC) thus represented a different holistic non-localized set of changes in all vertex positions. The first 50 shape PCs were used to construct faces that had a symmetric shape. Similarly, because face texture is also important for face perception, 50 texture PCs based on PCA of the RGB values of the faces were also used to represent faces. The resulting 300 faces were randomly generated from the 50 shape and 50 skin texture components with the constraint that all faces were set to be Caucasian. It is worth noting that each PC is a feature dimension of the face space.

Experimental procedure

We used a 1-back task for CelebA and FBI stimuli. In each trial, a single face was presented at the center of the screen for a fixed duration of 1s, with uniformly jittered inter trial interval (ITI) of 0.5-0.75s (**Fig. 1A**). Each image subtended a visual angle of approximately 10º. Patients pressed a button if the present face image was *identical* to the immediately previous image. 9% of trials were one-back repetitions. Each face was shown once unless repeated in one-back trials; and we excluded responses from one-back trials to have an equal number of responses for each face. This task kept patients attending to the faces, but avoided potential biases from focusing on a particular facial feature (e.g., compared to asking patients to judge a particular facial feature). The order of faces was randomized for each patient. This task procedure has been shown to be effective to study face representation in humans (3).

For FaceGen stimuli, patients performed two face judgment tasks. In each task, there was a judgment instruction, i.e., patients judged how trustworthy or how dominant a face was. We used a 1-4 scale: '1': not trustworthy / dominant at all, '2': somewhat trustworthy / dominant, '3': trustworthy / dominant, and '4': very trustworthy / dominant. Each image was presented for 1s at the center of the screen. One patient performed an additional passive-viewing task. We combined data from all tasks for analysis.

Stimuli were presented using MATLAB with the Psychtoolbox 3 (4) (http://psychtoolbox.org) (screen resolution: 1920 × 1080).

Feature extraction and construction of feature space

We used the well-known deep neural network (DNN) implementation based on the VGG-16 convolutional neural network (CNN) architecture (5) to extract features for each face image (see **Fig. S3A** for details). Fine-tuning was performed on the pre-trained VGG-Face deep model using all images of the 50 identities in the CelebA dataset (16-30 images for each identity). Features that differentiated identities (i.e., identity recognition) were extracted using this transferred model. The same network was also used in recent work (3) as the computational model for deep face feature extraction.

We subsequently applied a t-distributed stochastic neighbor embedding (t-SNE) method to convert high-dimensional features into a two-dimensional feature space. t-SNE is a variation of stochastic neighbor embedding (SNE) (6), a commonly used method for multiple class high-dimensional data visualization (7). We applied t-SNE for each layer, with the cost function parameter (Prep) of t-SNE, representing the perplexity of the conditional probability distribution induced by a Gaussian kernel, set individually for each layer. We implemented t-SNE in the MATLAB platform.

Notably, neither feature extraction nor construction of feature space utilized any information from neurons. Therefore, clustering of neurons in the feature space was not by construction.

Electrophysiology

We recorded from implanted depth electrodes in the amygdala and hippocampus from patients with pharmacologically intractable epilepsy. Target locations in the amygdala and hippocampus were verified using post-implantation CT. At each site, we recorded from eight 40 μm

microwires inserted into a clinical electrode as described previously (8, 9). Efforts were always made to avoid passing the electrode through a sulcus, and its attendant sulcal blood vessels, and thus the location varied but was always well within the body of the targeted area. Microwires projected medially out at the end of the depth electrode and examination of the microwires after removal suggests a spread of about 20-30 degrees. The amygdala electrodes were likely sampling neurons in the mid-medial part of the amygdala and the most likely microwire location is the basomedial nucleus or possibly the deepest part of the basolateral nucleus. Bipolar wide-band recordings (0.1-9kHz), using one of the eight microwires as reference, were sampled at 32kHz and stored continuously for off-line analysis with a Neuralynx system. The raw signal was filtered with zero-phase lag 300-3kHz bandpass filter and spikes were sorted using a semi-automatic template matching algorithm as described previously (10). Units were carefully isolated and recording and spike sorting quality were assessed quantitatively (**Fig. S1**).

Spikes

Only units with an average firing rate of at least 0.15Hz (entire task) were considered. Only single units were considered. Trials were aligned to stimulus onset. We used the mean firing rate in a time window 250ms to 1000ms after stimulus onset as the response to each face.

Selection of identity neurons

To select identity neurons, we first used a one-way ANOVA to identify neurons with significantly unequal response to different identities. We next imposed an additional criteria to define selective identities: the neural response was 2 standard deviations (SD) above the mean of neural responses from all identities. We refer to the neurons that encoded a single identity as single-identity (SI) neurons and we refer to the neurons that encoded multiple identities as multiple-identity (MI) neurons.

Selection of feature neurons

To select feature neurons, we first estimated a continuous spike density map in the feature space by smoothing the discrete firing rate map using a 2D Gaussian kernel (kernel size = feature dimension range * 0.2, SD = 4). We then estimated statistical significance for each pixel by permutation testing: in each of the 1000 runs, we randomly shuffled the labels of face examples. We calculated the p-value for each pixel by comparing the observed spike density value to those from the null distribution derived from permutation. We lastly selected the region with significant pixels (permutation $P < 0.05$, FDR corrected, cluster size > 3 * size of kernel dimension 1). We also applied a mask to exclude pixels from the edges and corners of the spike density map where there were no face examples because these regions were susceptible to false positives given our procedure. We selected feature neurons for each individual DNN layer. Because the distribution of faces was more sparse for FBI stimuli, we used a larger kernel (kernel size = feature value range * 0.3; cluster size > 1 * size of kernel dimension 1).

Identity selectivity index

To assess each neuron's selectivity to different identities, we defined an identity selectivity index as the *d'* between the most- and least-preferred identities (3):

$$\text{Identity Selectivity Index} = \frac{\mu_{\text{best}} - \mu_{\text{least}}}{\sqrt{\frac{1}{2}(\sigma^2_{\text{best}} + \sigma^2_{\text{least}})}}$$

where μ_{best} and μ_{worst} denote the mean firing rate for the most- and least-preferred identities, respectively, and σ^2_{best} and σ^2_{worst} denote the variance of firing rate for the most- and least-preferred identities, respectively.

Depth of selectivity (DOS) index

We quantified the DOS (11) for each neuron: $DOS = \frac{n - (\sum_{j=1}^{n} r_j)/r_{max}}{n-1}$, where n is the number of identities (n = 50), r_j is the mean firing rate to identity j, and r_{max} is the maximal mean firing rate across all identities. DOS varies from 0 to 1, with 0 indicating an equal response to all identities and 1 exclusive response to one identity, but not to any of the other identities. Thus, a DOS value of 1 is equal to maximal sparseness of identity coding.

Population decoding of face identities

We pooled all recorded neurons into a large pseudo-population (see (12, 13)). Firing rates were *z*-scored individually for each neuron to give equal weight to each unit regardless of firing rate. We used a maximal correlation coefficient classifier (MCC) as implemented in the MATLAB neural decoding toolbox (NDT) (14). The MCC estimates a mean template for each class *i* and assigns the class for test trial. We used 8-fold cross-validation, i.e., for each iteration 8 trials from each class were chosen from each neuron, among which 7 trials were used for training and the remaining 1 trial was used for testing. All possible train/test splits were tested and this process was repeated 50 times with different subsets of trials, resulting in a total of 400 tests to estimate the test performance. Spikes were counted in bins of 500ms size and advanced by a step size of 50ms. The first bin started −500ms relative to trial onset (bin center was thus 250ms before trial onset), and we tested 31 consecutive bins (the last bin was thus from 1000ms to 1500ms after trial onset). For each bin, a different classifier was trained/tested.

Web-association score

We employed a web-based association metric to study the relationship between different identities as done in a previous study (15). To estimate the degree of relationship between the 50 celebrity identities, we used an internet search engine (Google) and compared the number of hits to the joint searches with the number of hits to the individual searches. The rationale is that the

name of associated concepts will often appear together in web pages. The web-association score for each identity pair was calculated as $a_{ij} = \log_2\left(\frac{\text{hits}(\text{identity}_i \text{ AND } \text{identity}_j)}{\text{hits}(\text{identity}_i) \cdot \text{hits}(\text{identity}_j)}\right)$. Because we used celebrity faces, all identities were well searchable from the internet and could thus give a reasonable number of hits to calculate web-association values. We lastly normalized the web-association value using z-scoring.

Regression analyses

To identity neurons that encoded a linear combination of facial features, we employed both a partial least squares (PLS) regression with DNN feature maps (16, 17) and a linear regression with the two dimensions of the t-SNE feature space. For PLS, we used 25 components for each layer (16). For both approaches, we used a permutation test with 1000 runs to determine whether a neuron encoded a significant face model (i.e., the neuron encoded the dimensions of the face space). In each run, we randomly shuffled face labels and used 70% of faces as the training dataset. We used the training dataset to construct a model (i.e., deriving regression coefficients), predicted responses using this model for each face in the remaining 30% of faces (i.e., test dataset), and computed the Pearson correlation between the predicted and original response in the test dataset. The distribution of correlation coefficients computed *with* shuffling (i.e., null distribution) was eventually compared to the one *without* shuffling (i.e., observed response). If the correlation coefficient of the observed response was greater than 95% of the correlation coefficients from the null distribution, this face model was considered *significant*. This procedure has been shown to be very effective to select units with significant face models (18). The correlation coefficient could also indicate the model predictability and thus be compared between different neurons (**Fig. S9E, F, I, J**).

Table S1. List of patients.

ID	Age	Sex	Epilepsy diagnosis	Number of Amygdala Neurons					Number of Hippocampus Neurons				
				Total	Left	Right	SI	MI	Total	Left	Right	SI	MI
P6	33	F	Left posterior temporal/ parietal	9	9	0	0	0	31	31	0	5	3
				10	10	0	0	0	29	29	0	2	3
P7	28	F	Right mesial temporal	10	10	0	0	4	34	34	0	4	3
				6	6	0	0	0	21	19	2	0	0
				6	6	0	0	1	20	19	1	0	3
				2	2	0	0	1	31	26	5	3	1
P9	42	M	Left frontal	28	28	0	0	2	7	7	0	0	0
				29	29	0	0	0	7	7	0	0	1
				25	25	0	0	3	6	6	0	0	0
				23	23	0	0	1	3	3	0	0	0
P10	47	F	Right temporal	25	0	25	1	1	7	7	0	0	0
				24	0	24	0	0	27	27	0	1	1
P11	33	F	Right temporal	16	0	16	1	0	0	0	0	0	0
				14	0	14	0	0	0	0	0	0	0
				9	0	9	0	0	15	0	15	0	0
				6	0	6	0	0	10	0	10	0	1
Sum				242	148	94	2	13	248	215	33	15	16

Each row of neurons represents a separate recording session using the CelebA stimuli. Each session was recorded on a separate day. Total: all neurons recorded from an area. Left: neurons that were recorded from the left side of an area and had a firing rate greater than 0.15Hz. Right: neurons that were recorded from the right side of an area and had a firing rate greater than 0.15Hz. These neurons were included for further analysis. SI: identity neurons that encoded a single identity. MI: identity neurons that encoded multiple identities.

Supplementary Figure Legends

Fig. S1. Spike sorting and recording quality assessment. **(A)** Histogram of the number of units identified on each active wire (only wires with at least one unit identified are counted). The average yield per wire with at least one unit was 2.82±1.63 (mean±SD). **(B)** Histogram of mean firing rates. **(C)** Histogram of proportion of inter-spike intervals (ISIs) which are shorter than 3ms. The large majority of clusters had less than 0.5% of such short ISIs. **(D)** Histogram of the signal to noise ratio (SNR) of the mean waveform peak of each unit. **(E)** Histogram of the SNR of the entire waveform of all units. **(F)** Pairwise distance between all possible pairs of units on all wires where more than 1 cluster was isolated. Distances are expressed in units of standard deviation (SD) after normalizing the data such that the distribution of waveforms around their mean is equal to 1. **(G)** Isolation distance of all units for which this metric was defined (N = 490, median = 13.47). **(H)** Single-identity (SI) and multiple-identity (MI) neurons did not differ significantly in isolation distance ($t(34) = 0.99$, $P = 0.33$).

Fig. S2. More results for identity neurons. **(A)** Depth of selectivity (DOS) index. Identity neurons had a significantly higher DOS index than non-identity neurons. Error bars denote ±SEM across neurons. Asterisks indicate a significant difference using two-tailed unpaired *t*-test. **: $P < 0.01$. **(B, C)** Population decoding of face identity. **(B)** Decoding performance was primarily driven by identity neurons (red). As expected, the response of identity neurons was informative only about a small subset of identities. Shaded area denotes ±SEM across bootstraps. The horizontal dotted gray line indicates the chance level (2%). The top bars illustrate the time points with a significant above-chance decoding performance (bootstrap, $P < 0.05$, corrected by FDR for $Q < 0.05$). **(C)** MI neurons had a significantly better decoding performance than SI neurons because the encoding by MI neurons was less sparse. The top bar illustrates the time points with a significant difference between MI and SI neurons (bootstrap, $P < 0.05$, corrected by FDR for $Q < 0.05$). **(D)** Feature MI neurons did not differentiate familiar vs. unfamiliar selected identities (two-tailed paired *t*-test, $P > 0.05$). Error bars denote ±SEM across neurons.

Fig. S3. The deep neural network (DNN) used in this study. **(A)** Structure of the DNN. The convolutional neural network (CNN) consisted of a feature extraction section (13 convolutional layers) and a classification section (3 fully connected [FC] layers). The feature extraction section was consistent with the typical architecture of a CNN. A 3×3 filter with 1-pixel padding and 1-pixel stride was applied to each convolutional layer, which followed by a Batch Normalization (BatchNorm) and Rectified Linear Unit (ReLU) operation. Some of the convolutional layers were followed by five 2×2 max-pool operations with a stride of 2. There were 3 FC layers in each classification section: the first two had 4096 channels each, and the third performed an *n*-way classification. Each FC layer was followed by a ReLU and 50% dropout to avoid overfitting. A nonlinear Softmax operation was applied to the final output of VGG-16 network to make the classification prediction of 50 identities. **(B)** Feature visualization. **(C)** Feature correlation between faces. We calculated a correlation matrix of features between each face example, grouped by individual identity. It is worth noting that in earlier layers, features from face examples of the same identity were not highly correlated, suggesting that these face examples were not grouped together but distributed in the feature space (see also **Fig. S4**). **(D)** Pairwise distance between face examples in the full dimensional space was correlated with that in the t-SNE space. Shown are Pearson correlation coefficients *r*. Asterisk indicates a significant correlation in that layer ($P < 0.05$, Bonferroni-corrected for all layers).

Fig. S4. Projection of face examples onto feature space derived from each deep neural network (DNN) layer. Each color represents a different identity (names shown in the legend).

Fig. S5. More example feature neurons from the deep neural network (DNN) layer FC6. Legend conventions as in **Fig. 1**.

Fig. S6. Web-association matrix and correlation with deep neural network (DNN) features. **(A)** Web-association values between the 50 face identities. The color bar on the right shows the strength of association values in arbitrary units. **(B)** Correlation between the web-association score and DNN features. The web-association score for each identity pair was correlated with the Euclidean distance between their DNN features in each layer. Statistical significance was estimated by permutation testing: in each of the 1000 runs, labels of the web-association score were shuffled. P-values were calculated by comparing the observed correlation coefficient to those from the null distribution derived from the permutation.

Fig. S7. Projection of face examples onto feature space derived using uniform manifold approximation and projection (UMAP). Each color represents a different identity (names shown in the legend).

Fig. S8. More example feature neurons from deep neural network (DNN) layers Conv5_3 and Pool5. Legend conventions as in **Fig. 1**.

Fig. S9. Neurons showing a linear combination of facial features. **(A, B)** All neurons. **(C-F)** Identity neurons. **(G-J)** Feature neurons. **(A, C, E, G, I)** Partial least squares (PLS) regression with deep neural network (DNN) feature maps. **(B, D, F, H, J)** Linear regression with two t-SNE features (i.e., feature dimensions used to construct feature space for each layer). **(A-D, G, H)** The number of neurons showing a significant regression. Regression was performed individually for each DNN layer. Red dashed lines show the chance number of significant neurons (5% of all recorded neurons). Black bars show conditions with an above-chance number of significant neurons (binomial test: $P < 0.05$; after Bonferroni correction). **(E, F)** Comparison between identity and non-identity neurons. **(I, J)** Comparison between feature and non-feature neurons. Asterisks indicate a significant difference between identity vs. non-identity neurons or between

feature vs. non-feature neurons using two-tailed unpaired *t*-test. *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$, and ****: $P < 0.0001$. **(K)** Example neurons from **Fig. 1** showing that region-based feature coding could lead to a significant linear regression (shown by the plane). Although both approaches consistently showed that identity neurons and feature neurons had an above-chance number of selected neurons **(C, D, G, H)** as well as better model predictability than non-identity neurons **(E, F)** or non-feature neurons **(I, J)**, this was likely resulted from region-based feature coding, because an elevated response in one part of the feature space could drive the regression (see **(K)** for two examples).

Fig. S10. More results from the two additional experiments. **(A)** An example neuron demonstrating region-based feature coding in the FBI face space. **(B)** An example neuron demonstrating region-based feature coding in the FaceGen face space. **(C-E)** Common feature spaces for CelebA and FBI stimuli. The red outline delineates the tuning region of each example neuron. **(F)** Population results comparing neuronal response to FBI stimuli falling in vs. out of the tuning region. Error bars denote ±SEM across neurons. Asterisk indicates a significant difference between In vs. Out responses using paired *t*-test ($P < 0.05$).

References

1. Z. Liu, P. Luo, X. Wang, X. Tang, in Proceedings of International Conference on Computer Vision (ICCV). (2015).
2. N. N. Oosterhof, A. Todorov, The functional basis of face evaluation. *Proceedings of the National Academy of Sciences* **105**, 11087-11092 (2008).
3. S. Grossman *et al.*, Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. *Nature Communications* **10**, 4934 (2019).
4. D. H. Brainard, The Psychophysics Toolbox. *Spat Vis* **10**, 433-436 (1997).
5. O. M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition. (2015).
6. G. E. Hinton, S. T. Roweis, Stochastic neighbor embedding. *Advances in neural information processing systems*, 857-864 (2003).
7. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579-2605 (2008).
8. U. Rutishauser, A. N. Mamelak, E. M. Schuman, Single-Trial Learning of Novel Stimuli by Individual Neurons of the Human Hippocampus-Amygdala Complex. *Neuron* **49**, 805-813 (2006).
9. U. Rutishauser, I. B. Ross, A. N. Mamelak, E. M. Schuman, Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature* **464**, 903-907 (2010).
10. U. Rutishauser, E. M. Schuman, A. N. Mamelak, Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. *Journal of Neuroscience Methods* **154**, 204-224 (2006).
11. G. Rainer, W. F. Asaad, E. K. Miller, Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* **393**, 577-579 (1998).
12. U. Rutishauser *et al.*, Representation of retrieval confidence by single neurons in the human medial temporal lobe. *Nat Neurosci* **18**, 1041-1050 (2015).
13. S. Wang, A. N. Mamelak, R. Adolphs, U. Rutishauser, Abstract goal representation in visual search by neurons in the human pre-supplementary motor area. *Brain* **142**, 3530-3549 (2019).
14. E. Meyers, The Neural Decoding Toolbox. *Frontiers in Neuroinformatics* **7**, 1-12 (2013).
15. E. De Falco, M. J. Ison, I. Fried, R. Quian Quiroga, Long-term coding of personal and universal associations underlying the memory web in the human brain. *Nature Communications* **7**, 13408 (2016).
16. D. L. K. Yamins *et al.*, Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences* **111**, 8619 (2014).
17. C. R. Ponce *et al.*, Evolving Images for Visual Neurons Using a Deep Generative Network Reveals Coding Principles and Neuronal Preferences. *Cell* **177**, 999-1009.e1010 (2019).
18. L. Chang, D. Y. Tsao, The Code for Facial Identity in the Primate Brain. *Cell* **169**, 1013-1028.e1014 (2017).

Figure S1

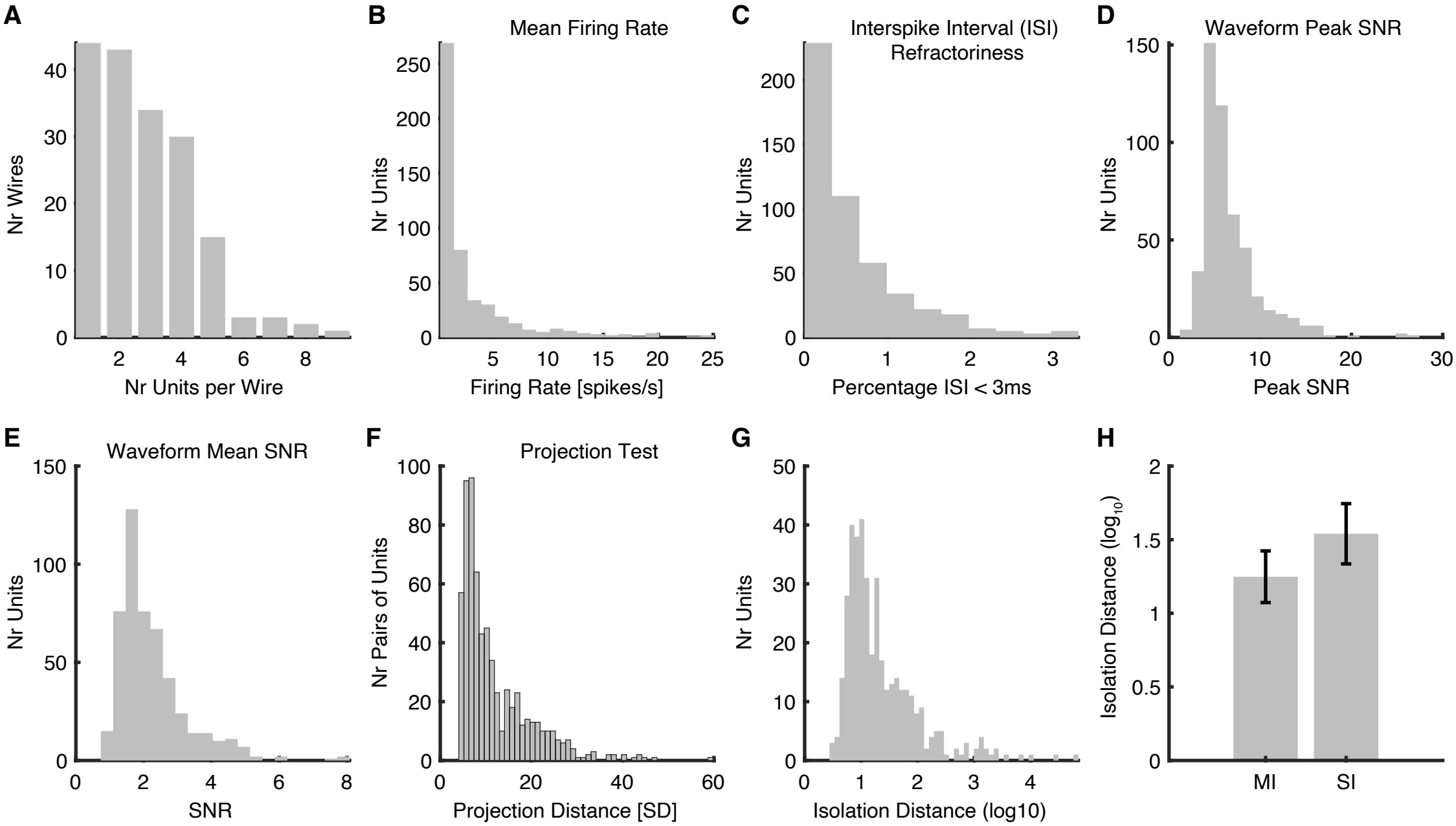
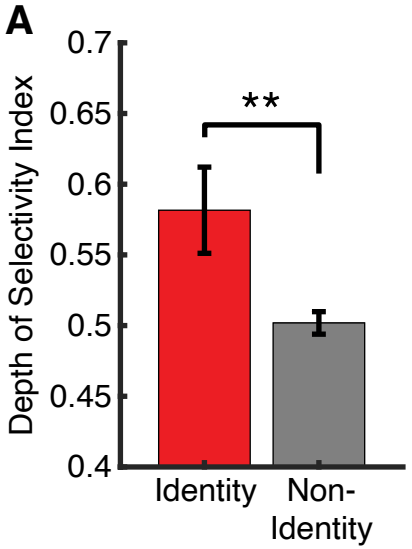
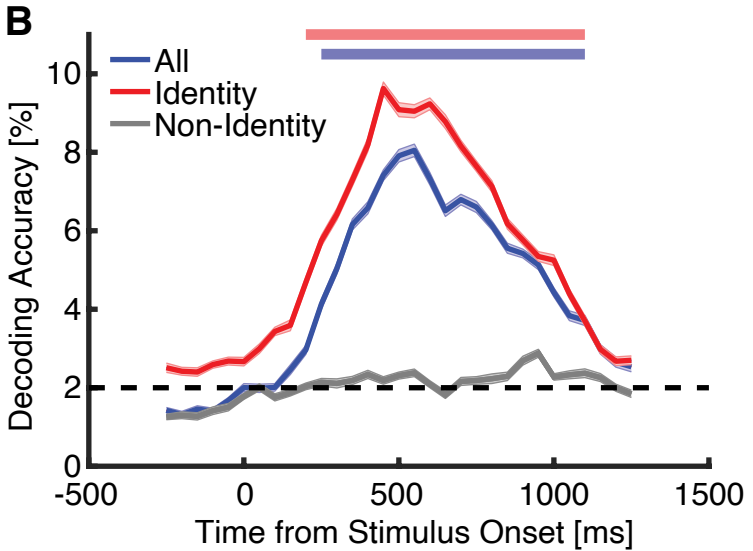
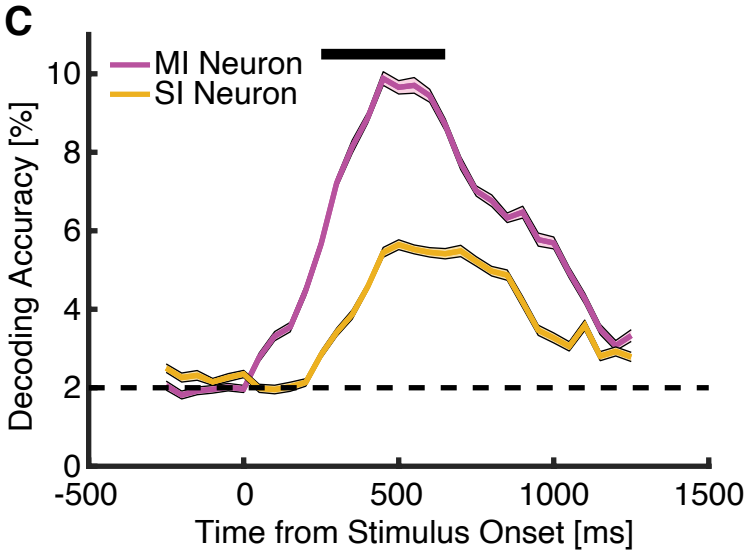

A
Nr Wires
Nr Units per Wire
B
Mean Firing Rate
Nr Units
Firing Rate [spikes/s]
C
Interspike Interval (ISI)
Refractoriness
Nr Units
Percentage ISI < 3ms
D
Waveform Peak SNR
Nr Units
Peak SNR
E
Waveform Mean SNR
Nr Units
SNR
F
Projection Test
Nr Pairs of Units
Projection Distance [SD]
G
Nr Units
Isolation Distance (log10)
H
Isolation Distance (log10)
MI
SI

Figure S2


A
Depth of Selectivity Index
0.7
0.65
0.6
0.55
0.5
0.45
0.4
**
Identity
Non-
Identity


B
All
Identity
Non-Identity
Decoding Accuracy [%]
10
8
6
4
2
0
-500
0
500
1000
1500
Time from Stimulus Onset [ms]


C
MI Neuron
SI Neuron
Decoding Accuracy [%]
10
8
6
4
2
0
-500
0
500
1000
1500
Time from Stimulus Onset [ms]

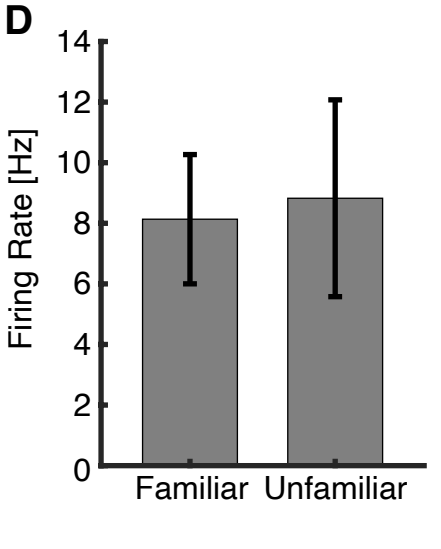
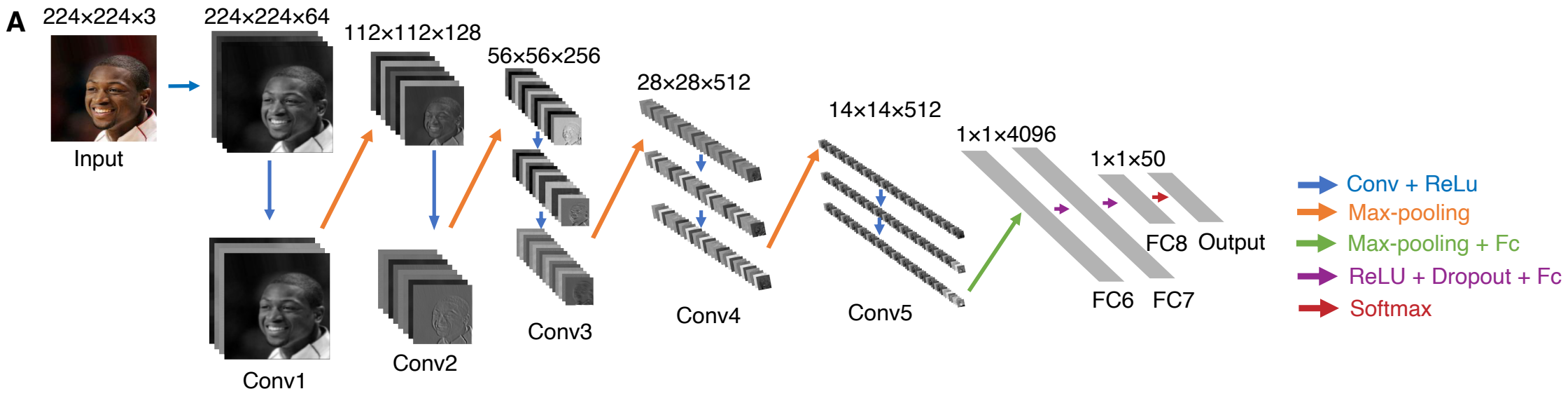
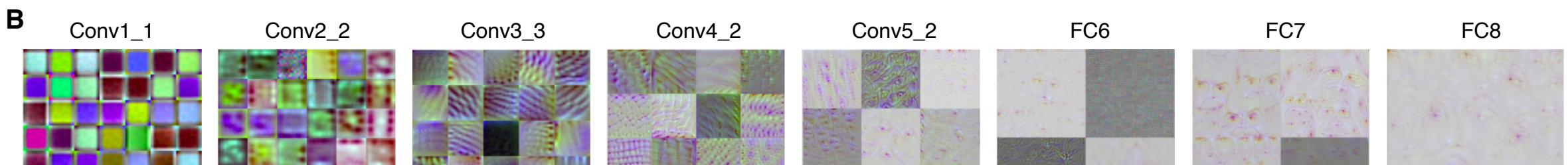
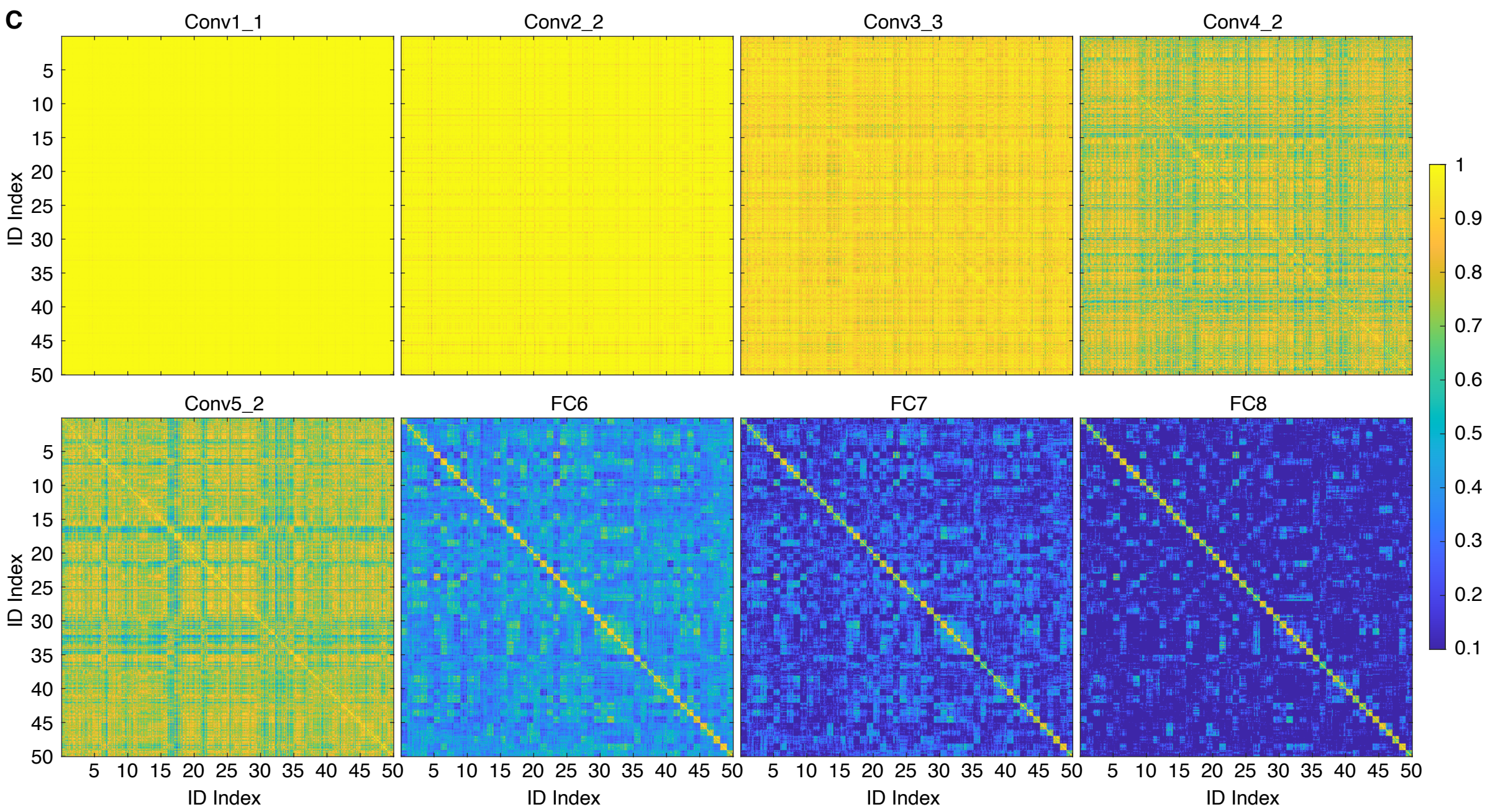

D
Firing Rate [Hz]
14
12
10
8
6
4
2
0
Familiar
Unfamiliar

Figure S3


A
224×224×3
224×224×64
112×112×128
56×56×256
28×28×512
14×14×512
1×1×4096
1×1×50
Input
Conv1
Conv2
Conv3
Conv4
Conv5
FC6
FC7
FC8
Output
Conv + ReLu
Max-pooling
Max-pooling + Fc
ReLU + Dropout + Fc
Softmax


B
Conv1_1
Conv2_2
Conv3_3
Conv4_2
Conv5_2
FC6
FC7
FC8


C
Conv1_1
Conv2_2
Conv3_3
Conv4_2
Conv5_2
FC6
FC7
FC8
ID Index
5
10
15
20
25
30
35
40
45
50
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

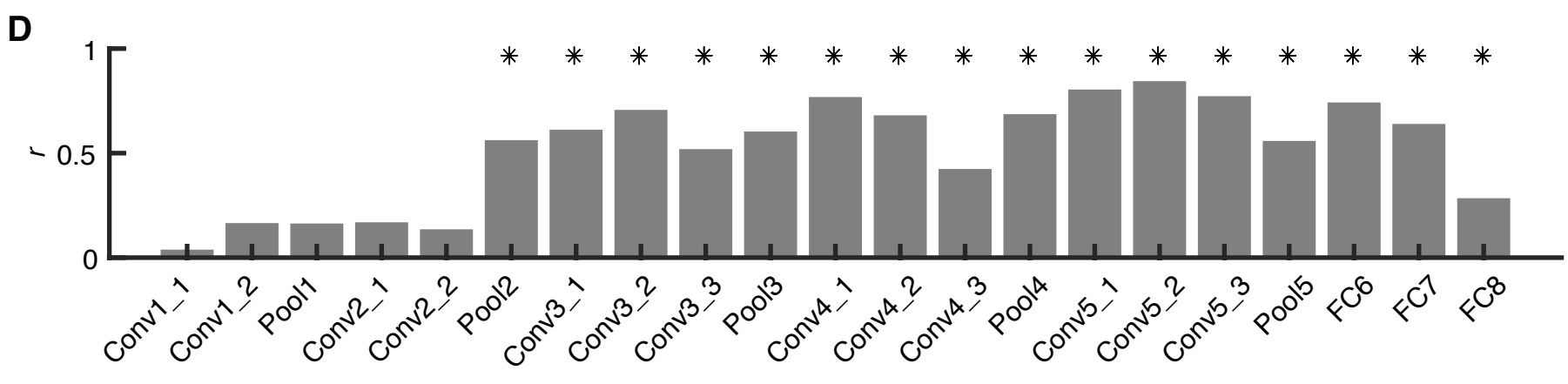

D
r
1
0.5
0
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8

Figure S4

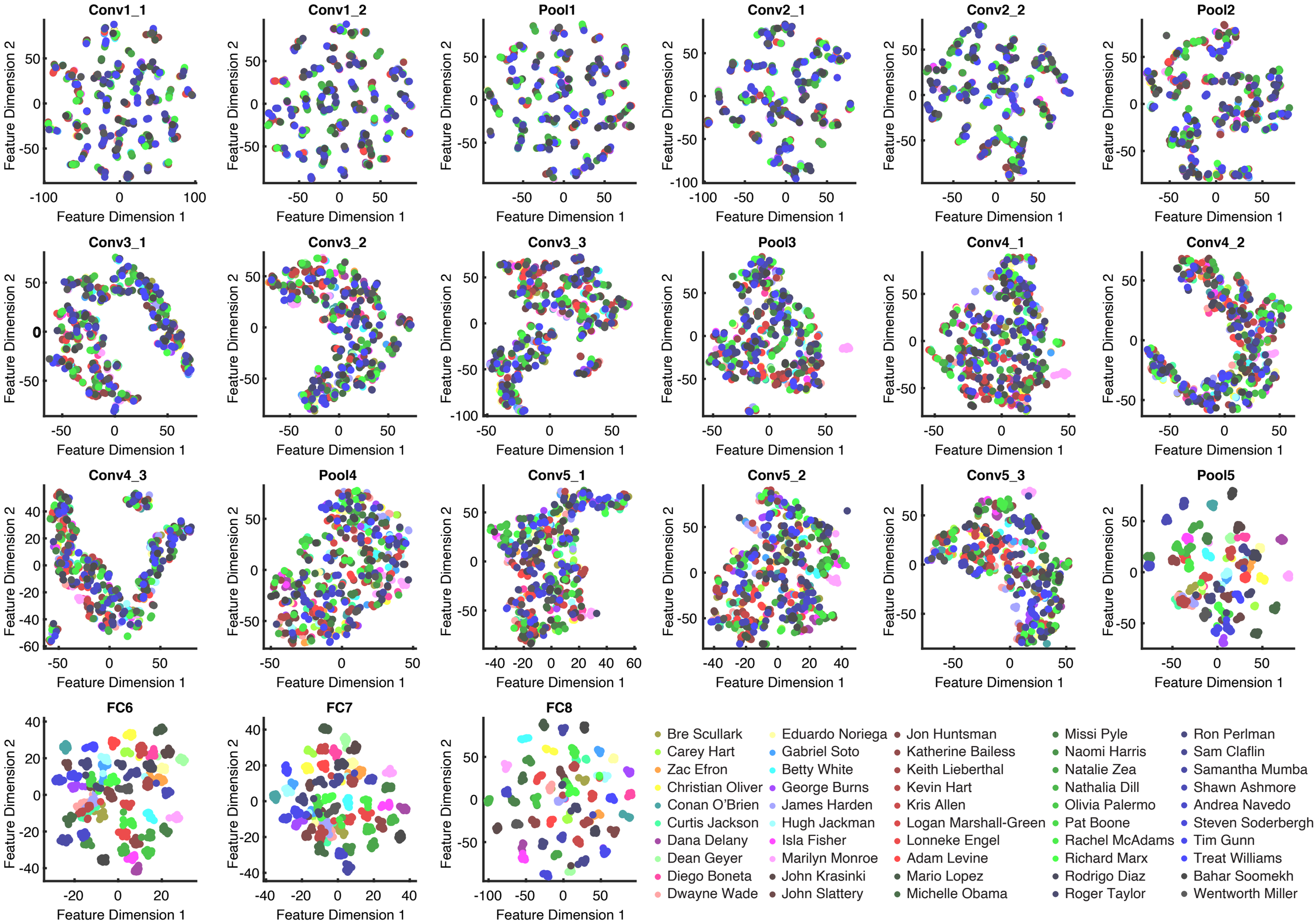

Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8
Feature Dimension 1
Feature Dimension 2
Bre Scullark
Carey Hart
Zac Efron
Christian Oliver
Conan O'Brien
Curtis Jackson
Dana Delany
Dean Geyer
Diego Boneta
Dwayne Wade
Eduardo Noriega
Gabriel Soto
Betty White
George Burns
James Harden
Hugh Jackman
Isla Fisher
Marilyn Monroe
John Krasinki
John Slattery
Jon Huntsman
Katherine Bailess
Keith Lieberthal
Kevin Hart
Kris Allen
Logan Marshall-Green
Lonneke Engel
Adam Levine
Mario Lopez
Michelle Obama
Missi Pyle
Naomi Harris
Natalie Zea
Nathalia Dill
Olivia Palermo
Pat Boone
Rachel McAdams
Richard Marx
Rodrigo Diaz
Roger Taylor
Ron Perlman
Sam Claflin
Samantha Mumba
Shawn Ashmore
Andrea Navedo
Steven Soderbergh
Tim Gunn
Treat Williams
Bahar Soomekh
Wentworth Miller

Figure S5

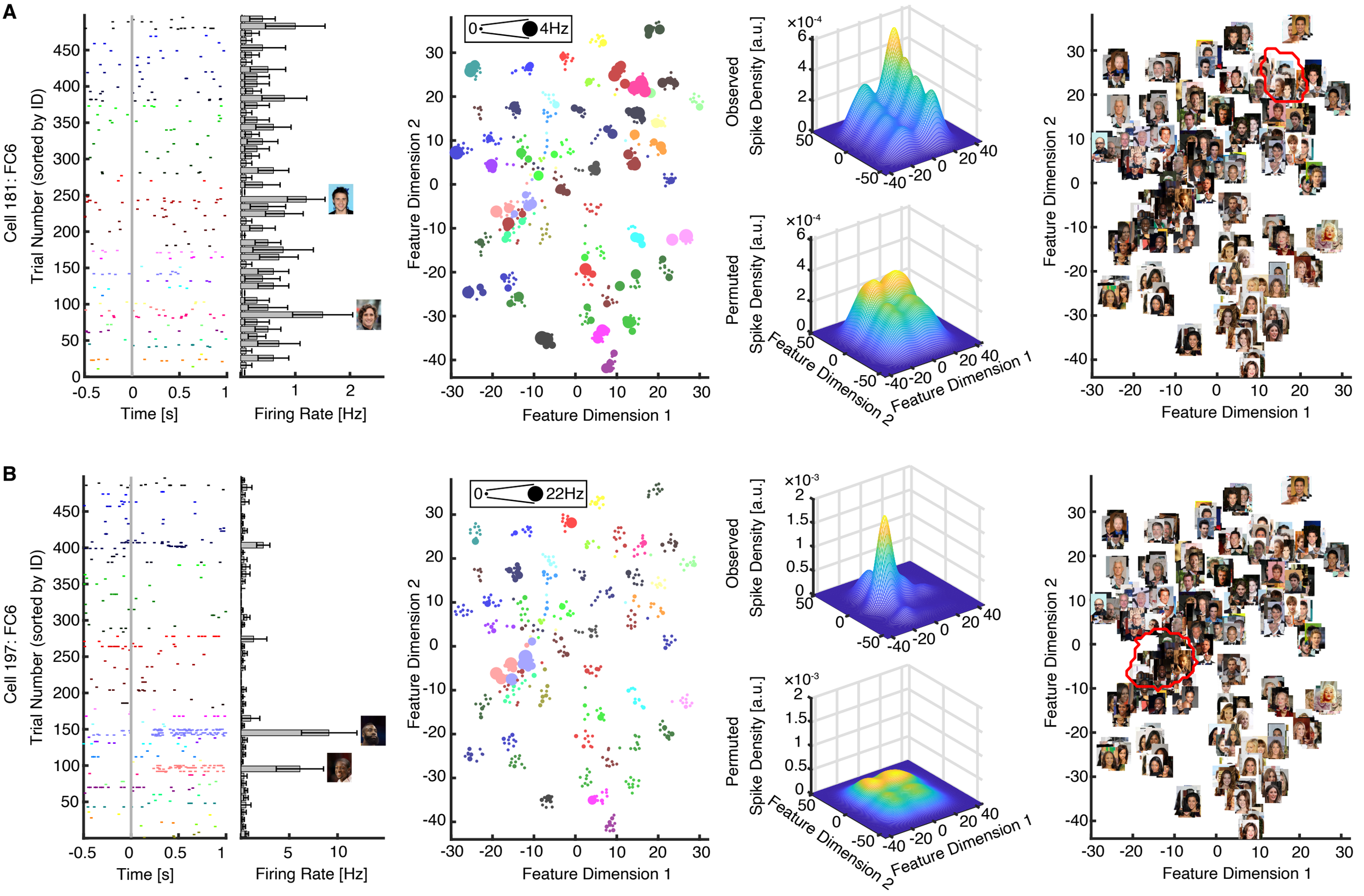
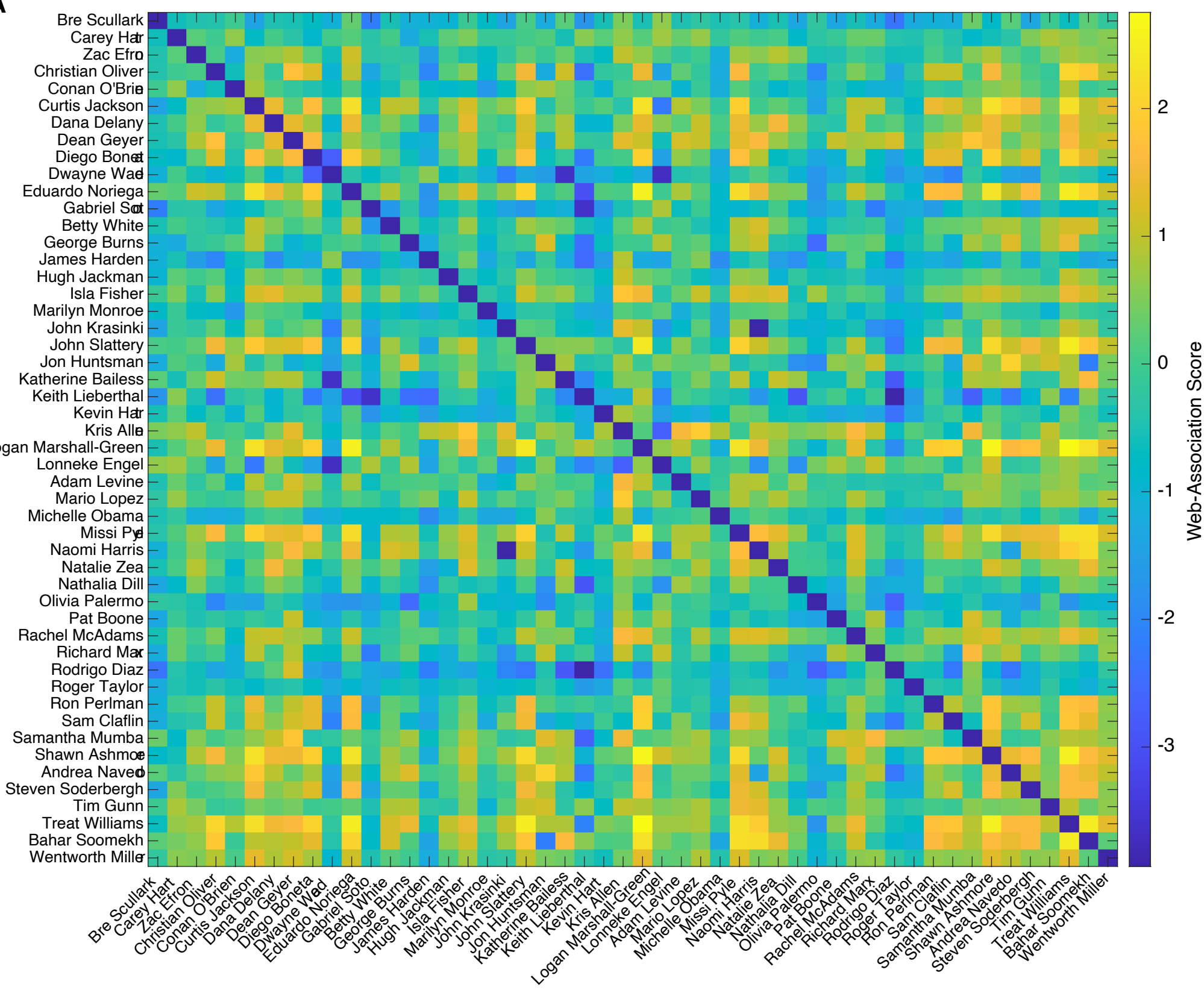

A
Cell 181: FC6
Trial Number (sorted by ID)
Time [s]
Firing Rate [Hz]
0
4Hz
Feature Dimension 2
Feature Dimension 1
Observed
Spike Density [a.u.]
Permuted
Spike Density [a.u.]
×10-4
B
Cell 197: FC6
22Hz
×10-3

Figure S6

A


Bre Scullark
Carey Hart
Zac Efron
Christian Oliver
Conan O'Brien
Curtis Jackson
Dana Delany
Dean Geyer
Diego Boneta
Dwayne Wade
Eduardo Noriega
Gabriel Soto
Betty White
George Burns
James Harden
Hugh Jackman
Isla Fisher
Marilyn Monroe
John Krasinki
John Slattery
Jon Huntsman
Katherine Bailess
Keith Lieberthal
Kevin Hart
Kris Allen
Logan Marshall-Green
Lonneke Engel
Adam Levine
Mario Lopez
Michelle Obama
Missi Pyle
Naomi Harris
Natalie Zea
Nathalia Dill
Olivia Palermo
Pat Boone
Rachel McAdams
Richard Marx
Rodrigo Diaz
Roger Taylor
Ron Perlman
Sam Claflin
Samantha Mumba
Shawn Ashmore
Andrea Navedo
Steven Soderbergh
Tim Gunn
Treat Williams
Bahar Soomekh
Wentworth Miller
2
1
0
-1
-2
-3
Web-Association Score

B

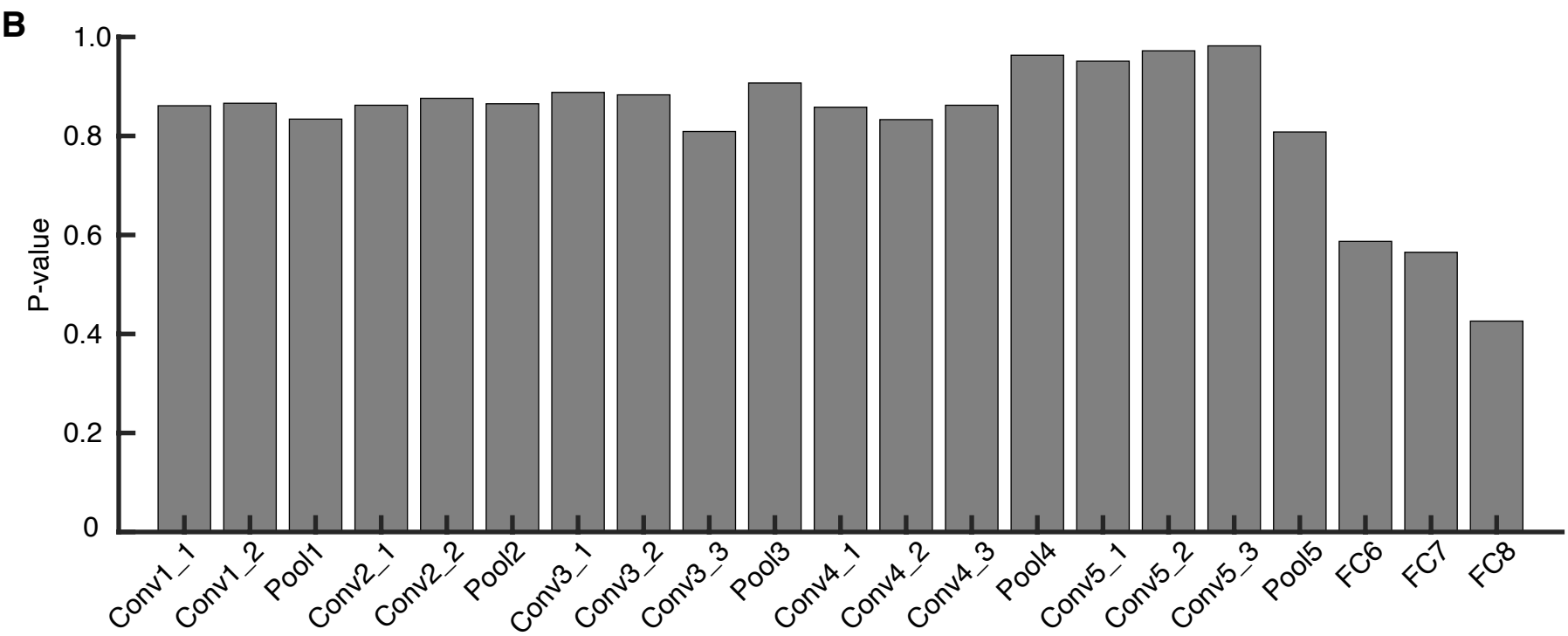

1.0
0.8
0.6
0.4
0.2
0
P-value
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8

Figure S7

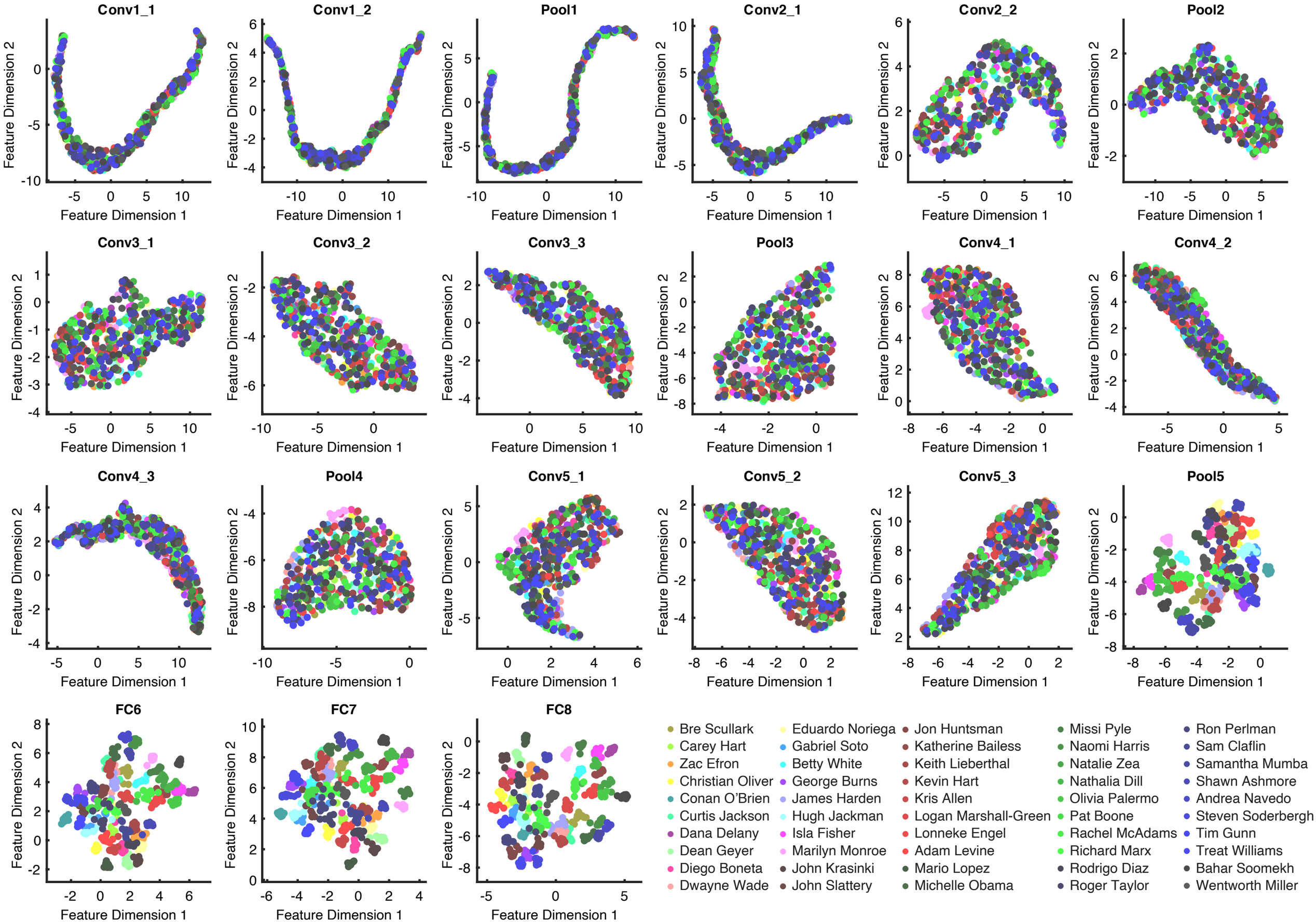

Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8
Feature Dimension 1
Feature Dimension 2
Bre Scullark
Carey Hart
Zac Efron
Christian Oliver
Conan O'Brien
Curtis Jackson
Dana Delany
Dean Geyer
Diego Boneta
Dwayne Wade
Eduardo Noriega
Gabriel Soto
Betty White
George Burns
James Harden
Hugh Jackman
Isla Fisher
Marilyn Monroe
John Krasinki
John Slattery
Jon Huntsman
Katherine Bailess
Keith Lieberthal
Kevin Hart
Kris Allen
Logan Marshall-Green
Lonneke Engel
Adam Levine
Mario Lopez
Michelle Obama
Missi Pyle
Naomi Harris
Natalie Zea
Nathalia Dill
Olivia Palermo
Pat Boone
Rachel McAdams
Richard Marx
Rodrigo Diaz
Roger Taylor
Ron Perlman
Sam Claflin
Samantha Mumba
Shawn Ashmore
Andrea Navedo
Steven Soderbergh
Tim Gunn
Treat Williams
Bahar Soomekh
Wentworth Miller

Figure S8

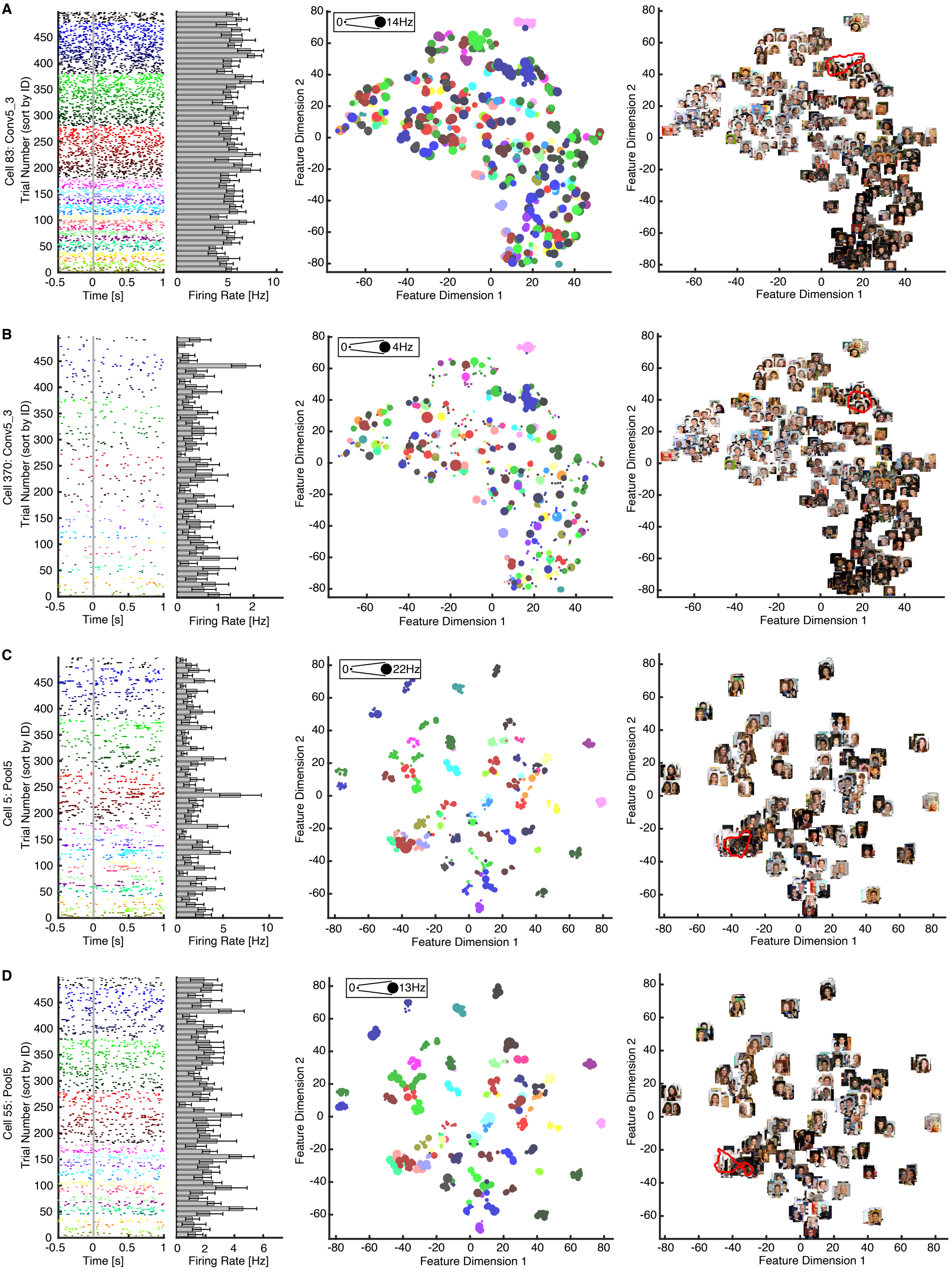
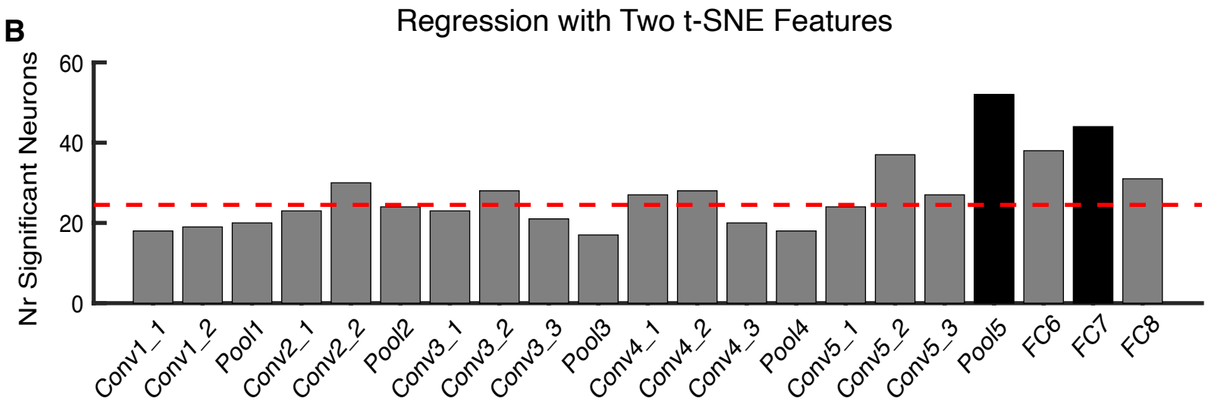
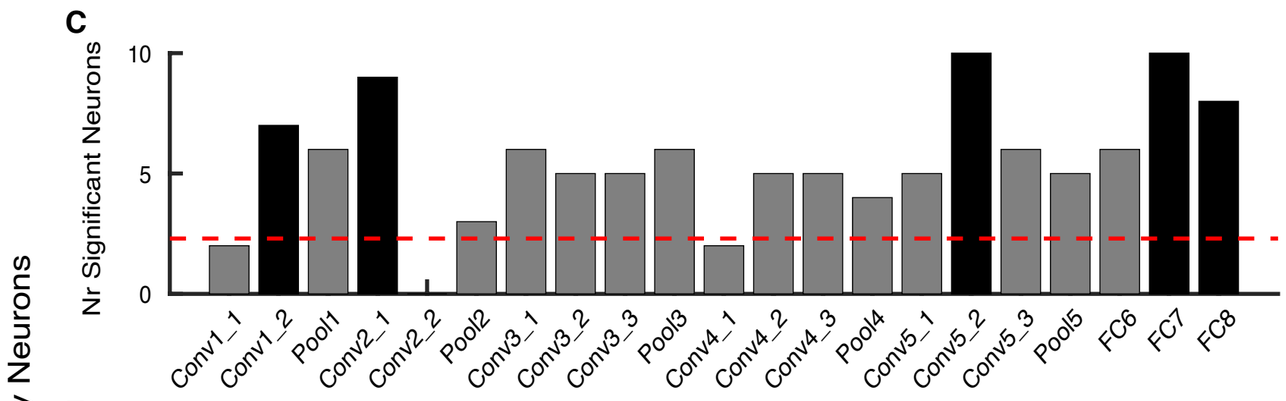
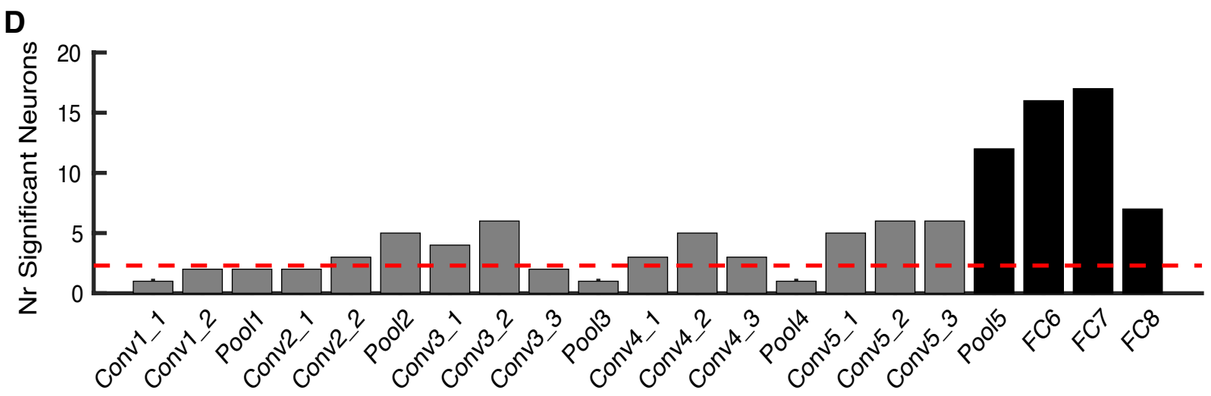

A
Cell 83: Conv5_3
Trial Number (sort by ID)
Time [s]
Firing Rate [Hz]
0
14Hz
Feature Dimension 2
Feature Dimension 1
B
Cell 370: Conv5_3
4Hz
C
Cell 5: Pool5
22Hz
D
Cell 55: Pool5
13Hz

Figure S9


A
Partial Least Squares (PLS) Regression with DNN Feature Maps
All Neurons
Nr Significant Neurons
0
10
20
30
40
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8

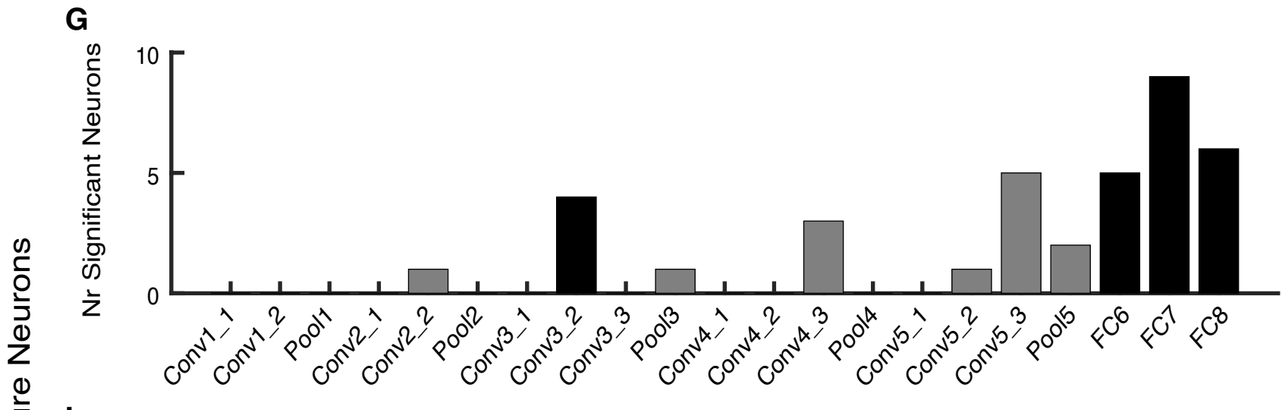

B
Regression with Two t-SNE Features
Nr Significant Neurons
0
20
40
60
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8

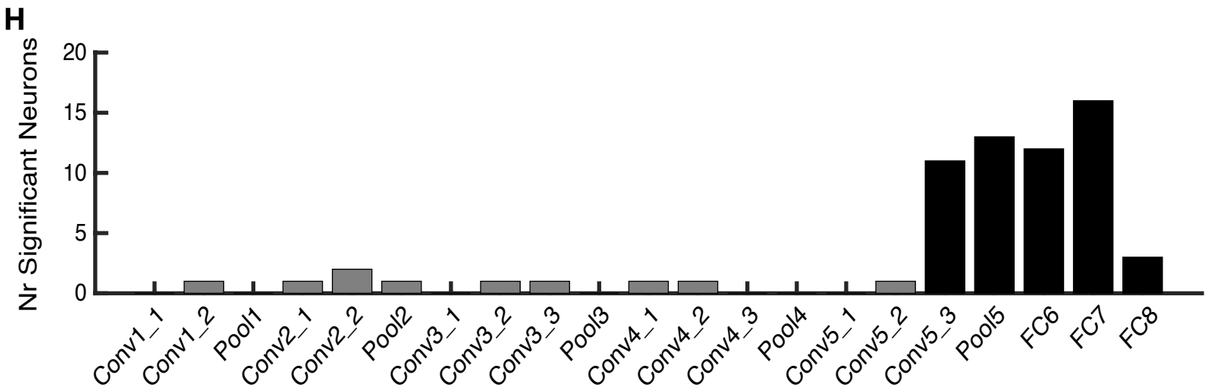

C
Nr Significant Neurons
0
5
10
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8

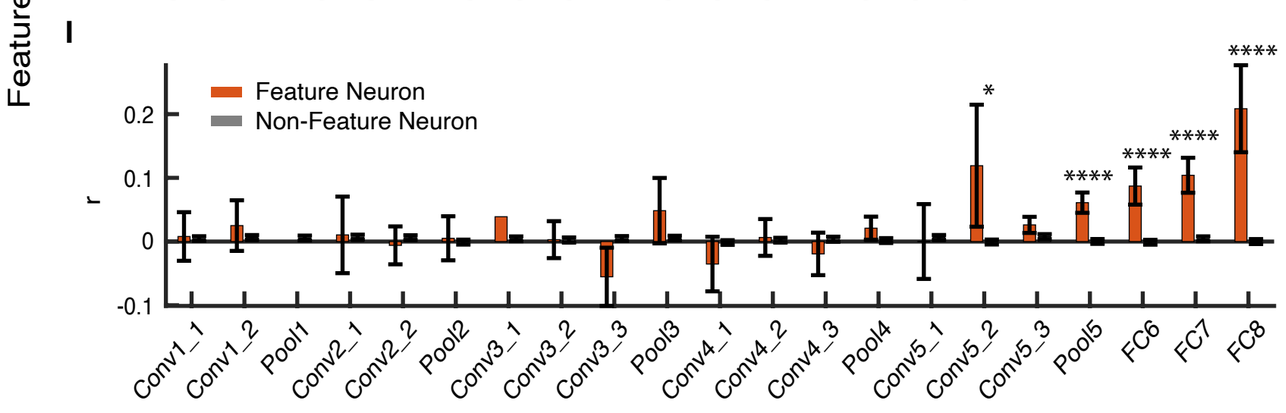

D
Nr Significant Neurons
0
5
10
15
20
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8

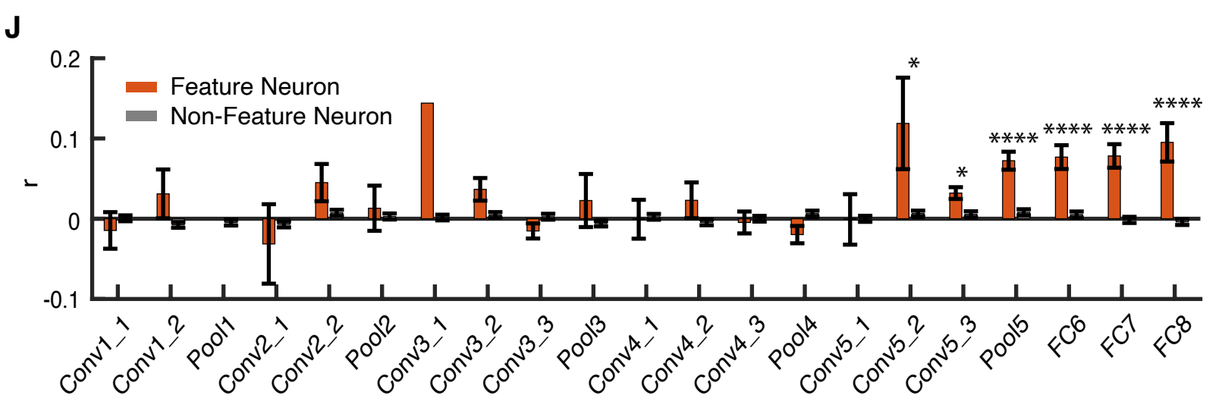
E
Identity Neurons
Identity Neuron
Non-Identity Neuron
r
-0.05
0
0.05
0.1
0.15
*
**
***
****
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8

F
Identity Neuron
Non-Identity Neuron
r
-0.05
0
0.05
*
**
***
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8


G
Nr Significant Neurons
0
5
10
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8


H
Nr Significant Neurons
0
5
10
15
20
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8


I
Feature Neurons
Feature Neuron
Non-Feature Neuron
r
-0.1
0
0.1
0.2
*
****
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8


J
Feature Neuron
Non-Feature Neuron
r
-0.1
0
0.1
0.2
*
****
Conv1_1
Conv1_2
Pool1
Conv2_1
Conv2_2
Pool2
Conv3_1
Conv3_2
Conv3_3
Pool3
Conv4_1
Conv4_2
Conv4_3
Pool4
Conv5_1
Conv5_2
Conv5_3
Pool5
FC6
FC7
FC8

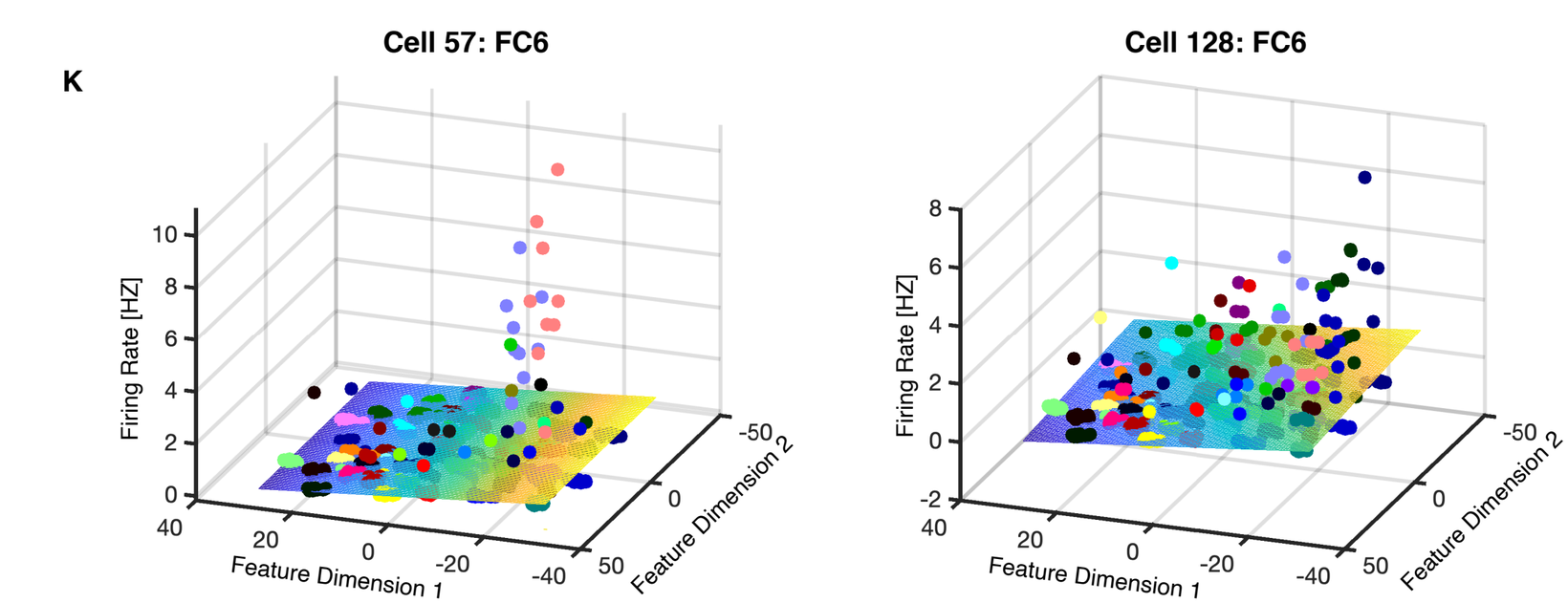
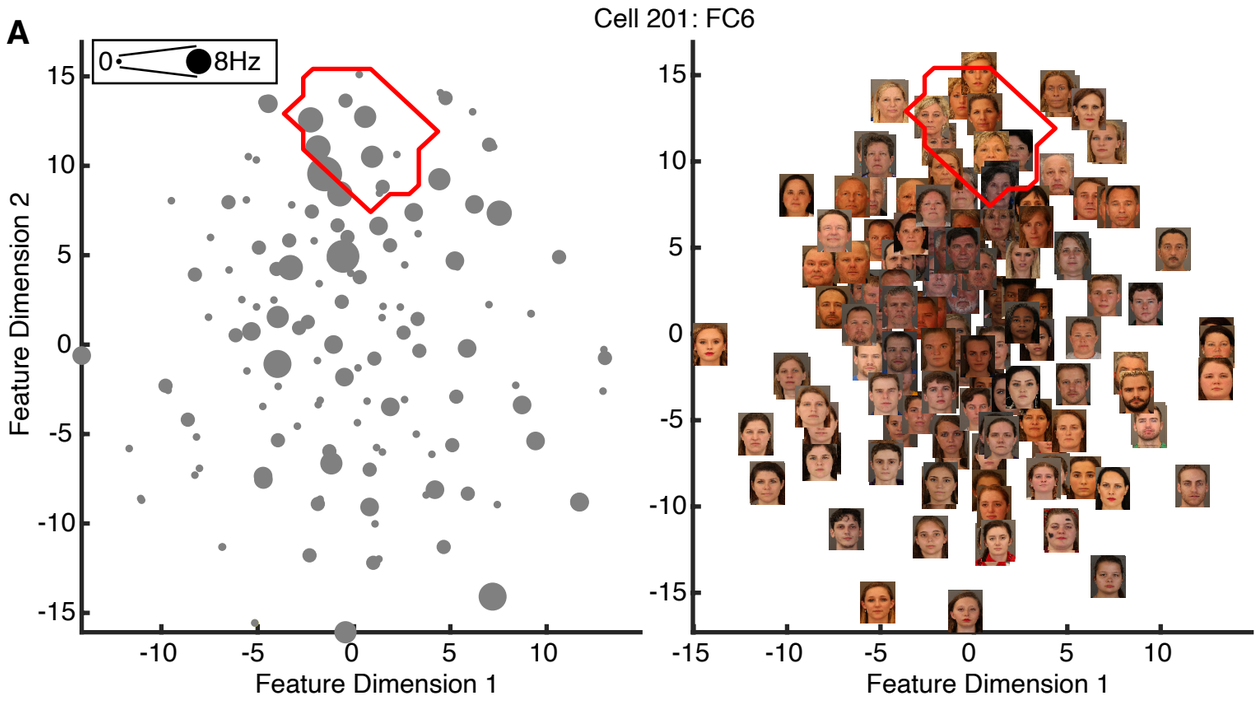
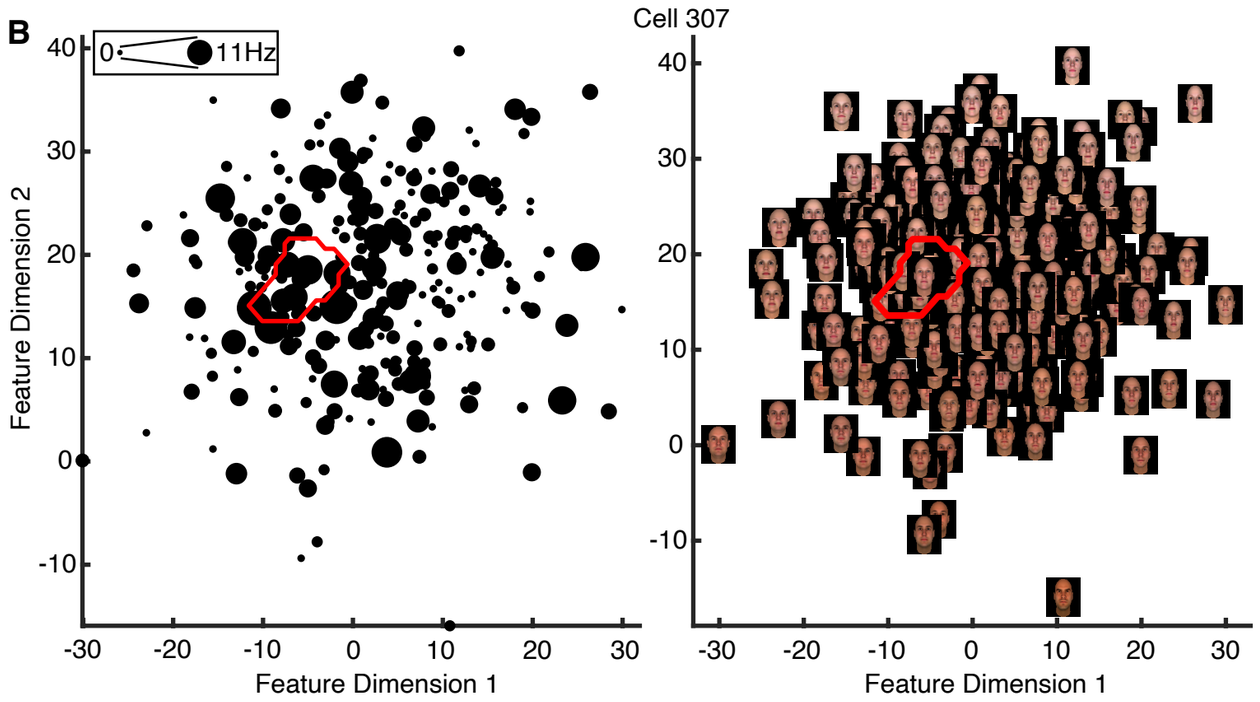
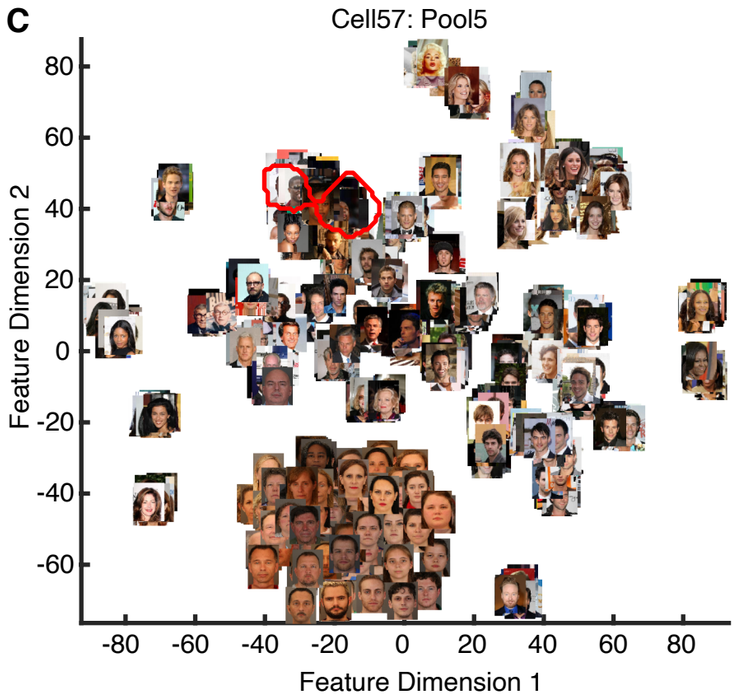
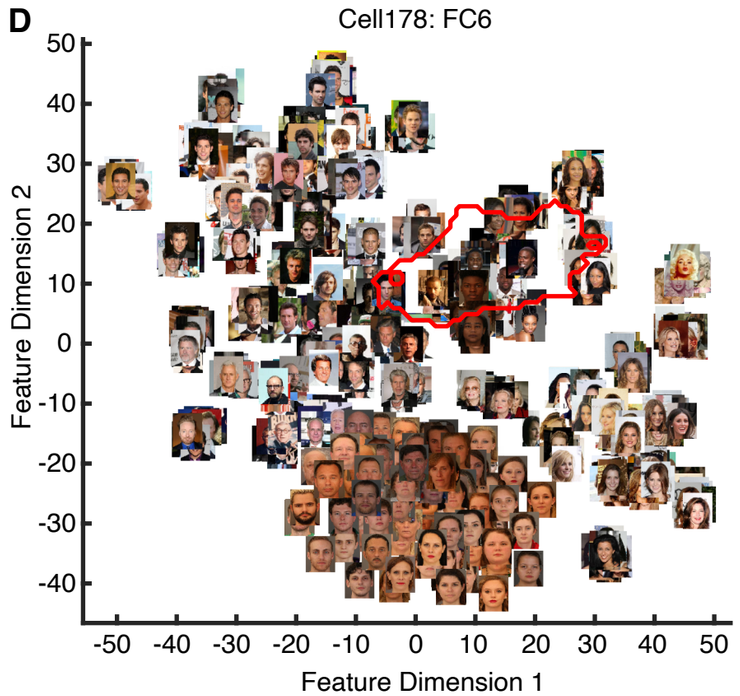

K
Cell 57: FC6
Cell 128: FC6
Firing Rate [HZ]
Feature Dimension 1
Feature Dimension 2
40
20
0
-20
-40
50
0
-50

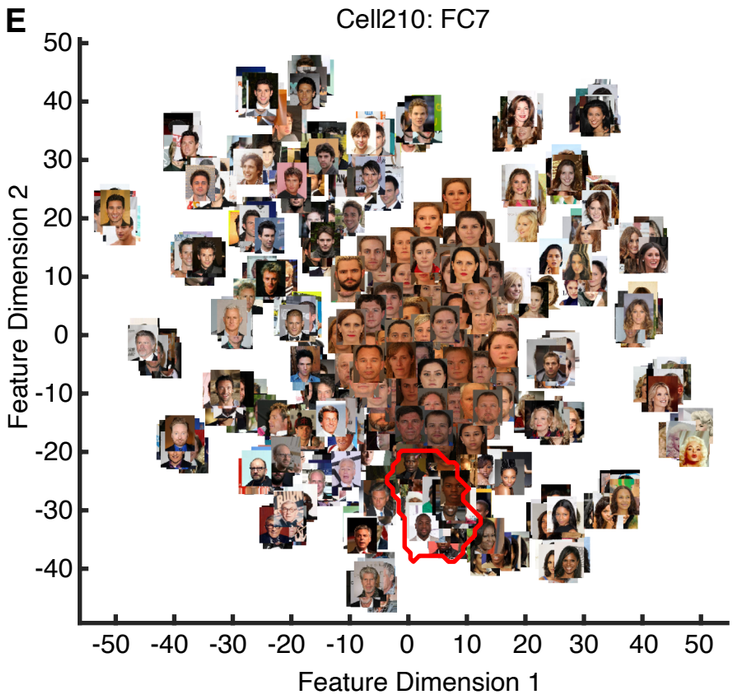
Figure S10

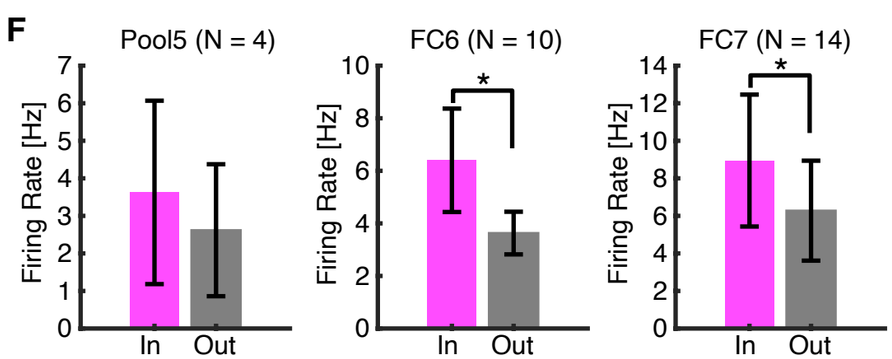

A
Cell 201: FC6
0
8Hz
Feature Dimension 2
Feature Dimension 1


B
Cell 307
0
11Hz
Feature Dimension 2
Feature Dimension 1


C
Cell57: Pool5
Feature Dimension 2
Feature Dimension 1


D
Cell178: FC6
Feature Dimension 2
Feature Dimension 1


E
Cell210: FC7
Feature Dimension 2
Feature Dimension 1


F
Pool5 (N = 4)
FC6 (N = 10)
FC7 (N = 14)
Firing Rate [Hz]
In
Out
*